

大自由度ランダム RNN の動的平均場理論とその周辺

高須 正太郎 *

2024 年 3 月 8 日

1 モデル設定

Sompolinsky, Crisanti, Sommers によって提案された以下のランダム RNN モデルを考える [1, 2] :

$$(1 + \partial_t)x_i(t) = \sum_j^N J_{ij}\phi_j(t) + \xi_i(t) \quad (1)$$

ここで活性化関数を $\phi(x)$ として、 $\phi_j(t) \equiv \phi(x_j(t))$ と略記している。 ϕ はシグモイド関数、つまり単調増加な奇関数で、 $\phi'(0) = 1$ 、 $\phi''(x) < 0$ ($x \geq 0$)、 $\lim_{x \rightarrow \pm\infty} \phi(x) = \pm 1$ を満たす関数である。 N は RNN のニューロン数を表し、以下の解析では $N \gg 1$ とする。 $\xi_i(t)$ は Gaussian white noise であり、 $\langle \xi_i(t) \rangle = 0$ 、 $\langle \xi_i(t)\xi_j(s) \rangle = D\delta_{ij}\delta(t-s)$ とする。 J_{ij} はニューロン j からニューロン i への結合重みを表し、Gauss 分布 $\mathcal{N}(\bar{g}/N, g^2/N)$ からの i.i.d サンプルであるとする。なお、 J_{ij} の平均が非ゼロであり、パラメーター \bar{g} にコントロールされている点は、Sompolinsky らによる元々のモデル ($\bar{g} = 0$) と異なる。Gauss 分布の平均・分散が $O(1/N)$ であるのは、Eq.1 の右辺第 1 項の平均・分散が $O(1)$ となることを保証するためである。

2 自明固定点の安定性

動的平均場理論に入る前に、この RNN の固定点、特に自明固定点の安定性を見ておく。本章では、 $D = 0$ 、つまり入力がない場合のみを考える。このとき自明固定点は $\mathbf{x} = \mathbf{0}$ であり、固定点の安定性は Jacobian $-I + J$ の固有値が複素平面の $\Re z = 0$ をまたぐかどうかで決まる。つまり、 J の固有値が $\Re z > 1$ にあるかどうかを調べれば良い。各成分が $\mathcal{N}(0, g^2/N)$ からの i.i.d サンプルであるランダム行列 J_0 を用いて、 $J = \frac{\bar{g}}{N}\mathbf{1}\mathbf{1}^\top + J_0$ と表せる。circular law より J_0 の固有値は、原点を中心とする半径 g の円内に均一に分布する。 J は J_0 に 1-rank 行列 $\frac{\bar{g}}{N}\mathbf{1}\mathbf{1}^\top$ の摂動が加わったものであるが、このような行列の固有値分布は、前述の半径 g の円と 1 点 $z = \bar{g}$ になることが知られている [3, 4]。したがって、 $g < 1$ かつ $\bar{g} < 1$ が自明固定点の安定条件であり、いずれかが破られると不安定化する。

$g < 1$ のもとの、 \bar{g} が分岐点 $\bar{g} = 1$ を超えるシナリオを考える。すると、固定点 $\mathbf{x} = \mathbf{0}$ が不安定化し、新たな固定点 \mathbf{x} に転移する (下図の左図)。このとき素子間相関 $\langle \phi_i(t)\phi_j(t) \rangle_t$ は非ゼロになるので、各素子が強い相関を持った活動を示す regime になる。これは神経雪崩 (neural avalanches) で見られる活動であり、 \bar{g} を神経雪崩への転移を司るパラメータとみなすことができる [5]。

一方、 $\bar{g} < 1$ のもとの g が 1 を超えるシナリオでは、RNN はカオスダイナミクスを示す (下図の右図)。このとき、素子間相関はほぼ 0 になるため、各素子は互いにほぼ独立に活動している点が神経雪崩とは異なる。 g は chaos への転移を司るパラメータとなる。

* 京都大学大学院 情報学研究所 先端数理科学専攻 博士課程
website: <https://shtakasu.github.io/>

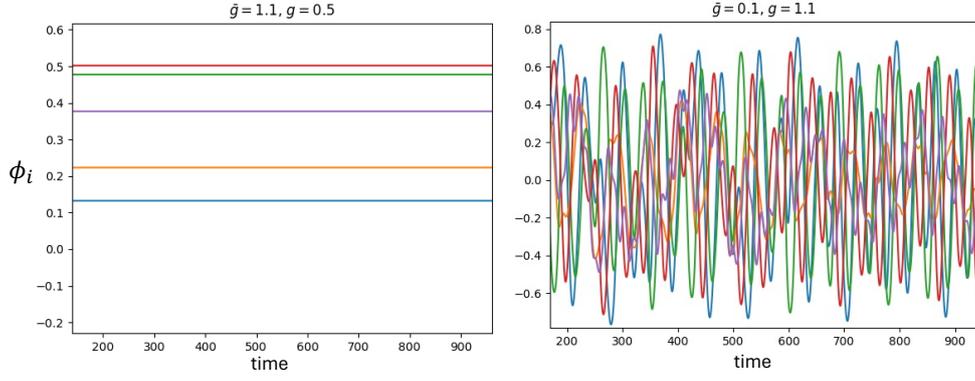


図1 左図：neural avalanche 転移、右図：chaos 転移

3 動的平均場理論

RNN のマクロなダイナミクスを理解すべく、 $R(t) \equiv \frac{\bar{g}}{N} \sum_i^N \phi_i(t)$ 、 $Q(t, s) \equiv \frac{g^2}{N} \sum_i^N \phi_i(t)\phi_i(s)$ というオーダーパラメータを導入し、この統計量を調べる。

(1) 生成汎関数の経路積分表示

Eq.1 の trajectory $\mathbf{x}(t)$ を引数に取る汎関数 $G[\mathbf{x}]$ の試行平均は、確率微分方程式の Martin - Siggia - Rose-De Dominicis - Janssen 経路積分形式を用いて、

$$\begin{aligned} \langle G[\mathbf{x}] \rangle_{\mathbf{x}|J} &= \int_{\mathbf{x}, \tilde{\mathbf{x}}} G[\mathbf{x}] \exp \left(S_0[\mathbf{x}, \tilde{\mathbf{x}}] - \tilde{\mathbf{x}}^\top J \mathbf{x} \right) \\ S_0[\mathbf{x}, \tilde{\mathbf{x}}] &:= \tilde{\mathbf{x}}^\top (\partial_t + 1) \mathbf{x} + \frac{D}{2} \tilde{\mathbf{x}}^\top \tilde{\mathbf{x}} \end{aligned}$$

となる（導出は文献 [6] 第 7 章を参照）。ここで、 $\mathbf{a}^\top \mathbf{b} := \sum_i^N \int a_i(t) b_i(t) dt$ 、 $\mathbf{a}^\top M \mathbf{b} := \sum_{i,j}^N \int a_i(t) M_{ij} b_j(t) dt$ という表記を用いた。また経路積分の測度 $\mathcal{D}\mathbf{x}$ 、 $\mathcal{D}\tilde{\mathbf{x}}$ は省略している。 $S_0[\mathbf{x}, \tilde{\mathbf{x}}]$ は相互作用 J が無い、自由場の action を表している。 $G[\mathbf{x}] = \exp(\mathbf{j}^\top \mathbf{x})$ をとれば、 $\langle \exp(\mathbf{j}^\top \mathbf{x}) \rangle_{\mathbf{x}|J}$ は Eq.1 のモーメント生成汎関数になる。

興味のあるオーダーパラメーター $R(t)$ 、 $Q(t, s)$ はいずれも自己平均性が期待できるので、 $\langle G[\mathbf{x}] \rangle_{\mathbf{x}|J}$ を J に関して平均（クエンチ平均）をとると、

$$\begin{aligned} \langle \langle G[\mathbf{x}] \rangle_{\mathbf{x}|J} \rangle_J &= \int_{\mathbf{x}, \tilde{\mathbf{x}}} G[\mathbf{x}] e^{S_0[\mathbf{x}, \tilde{\mathbf{x}}]} \langle e^{-\tilde{\mathbf{x}}^\top J \mathbf{x}} \rangle_J \\ &= \int_{\mathbf{x}, \tilde{\mathbf{x}}} G[\mathbf{x}] \exp \left(S_0[\mathbf{x}, \tilde{\mathbf{x}}] - \frac{\bar{g}}{N} \sum_{i,j} \tilde{x}_i^\top \phi_j + \frac{1}{2} \frac{g^2}{N} \sum_{i,j} (\tilde{x}_i^\top \phi_j)^2 \right) \\ &= \int_{\mathbf{x}, \tilde{\mathbf{x}}} G[\mathbf{x}] \exp \left(S_0[\mathbf{x}, \tilde{\mathbf{x}}] - \frac{\bar{g}}{N} \sum_{i,j} \tilde{x}_i^\top \phi_j + \frac{1}{2} \frac{g^2}{N} \int dt ds \sum_i \tilde{x}_i(t) \tilde{x}_i(s) \sum_j \phi_j(t) \phi_j(s) \right) \end{aligned}$$

となる。オーダーパラメータの定義を上式に課すために、デルタ関数のフーリエ変換表示を用いた恒等式

$$\begin{aligned} 1 &\propto \int_R \delta \left[NR(t) - \bar{g} \sum_i \phi_i(t) \right] \propto \int_{R, \tilde{R}} \exp \left(NR^\top \tilde{R} - \bar{g} \sum_i \phi_i^\top \tilde{R} \right) \\ 1 &\propto \int_Q \delta \left[NQ(t, s) - g^2 \sum_i \phi_i(t) \phi_i(s) \right] \propto \int_{Q, \tilde{Q}} \exp \left(NQ^\top \tilde{Q} - g^2 \sum_i \phi_i^\top \tilde{Q} \phi_i \right) \end{aligned}$$

を挿入することによって、

$$\begin{aligned} \langle \langle G[\mathbf{x}] \rangle_{\mathbf{x}|J} \rangle_J &= \frac{1}{A} \int_y \int_{\mathbf{x}, \tilde{\mathbf{x}}} G[\mathbf{x}] \exp \left(N \left(R^\top \tilde{R} + Q^\top \tilde{Q} \right) + \sum_i \left(S_0[x_i, \tilde{x}_i] - \tilde{x}_i^\top R + \frac{1}{2} \tilde{x}_i^\top Q \tilde{x}_i - \bar{g} \phi_i^\top \tilde{R} - g^2 \phi_i^\top \tilde{Q} \phi_i \right) \right) \\ A &:= \int_y \int_{\mathbf{x}, \tilde{\mathbf{x}}} \exp \left(N \left(R^\top \tilde{R} + Q^\top \tilde{Q} \right) + \sum_i \left(S_0[x_i, \tilde{x}_i] - \tilde{x}_i^\top R + \frac{1}{2} \tilde{x}_i^\top Q \tilde{x}_i - \bar{g} \phi_i^\top \tilde{R} - g^2 \phi_i^\top \tilde{Q} \phi_i \right) \right) \end{aligned} \quad (2)$$

が得られる。ここで $y := \{R, \tilde{R}, Q, \tilde{Q}\}$ および、 $Q^\top \tilde{Q} := \int Q(t, s) \tilde{Q}(t, s) ds dt$ という表記を用いた。Eq.2 では異なるニューロン同士の積がないことから、各ニューロンは、共通の場 y のみを介して他のニューロンと相互作用するという物理的描像が浮かび上がってくる。一方、この場 y はニューロンの活動によって定まる。したがって、self-consistent な関係式が導出できると予想できる。この関係式がいわゆる動的平均場方程式であり、この一連の流れが RNN の動的平均場理論の本質である。

オーダーパラメータ R, Q に興味があるので、補助変数 $j = \{j_1, j_2, j_3, j_4\}$ を導入し、Eq.2 において $G[\mathbf{x}] = \exp(j^\top y)$ とする。ここで、 $j^\top y := \int_t j_1(t) R(t) + \int_t j_2(t) \tilde{R}(t) + \int_{t,s} j_3(t, s) Q(t, s) + \int_{t,s} j_4(t, s) \tilde{Q}(t, s)$ と表している。すると Eq.2 は、

$$\begin{aligned} Z[j] &:= \langle \langle \exp(j^\top y) \rangle_{\mathbf{x}|J} \rangle_J \\ &= \frac{1}{A} \int_y \int_{\mathbf{x}, \tilde{\mathbf{x}}} \exp \left(N \left(R^\top \tilde{R} + Q^\top \tilde{Q} \right) + j^\top y + \sum_i \left(S_0[x_i, \tilde{x}_i] - \tilde{x}_i^\top R + \frac{1}{2} \tilde{x}_i^\top Q \tilde{x}_i - \bar{g} \phi_i^\top \tilde{R} - g^2 \phi_i^\top \tilde{Q} \phi_i \right) \right) \\ &= \frac{1}{A} \int_y \exp \left(N \left(R^\top \tilde{R} + Q^\top \tilde{Q} + \mathcal{W}[y] \right) + j^\top y \right) \\ \mathcal{W}[y] &:= \ln \int_{\mathbf{x}, \tilde{\mathbf{x}}} \exp \left(S_0[x, \tilde{x}] - \tilde{x}^\top R + \frac{1}{2} \tilde{x}^\top Q \tilde{x} - \bar{g} \phi^\top \tilde{R} - g^2 \phi^\top \tilde{Q} \phi \right) \\ A &:= \int_y \exp \left(N \left(R^\top \tilde{R} + Q^\top \tilde{Q} + \mathcal{W}[y] \right) \right) \end{aligned} \quad (3)$$

となる。オーダーパラメータに関するモーメント生成汎関数 $Z[j]$ の対数は、キュムラント生成汎関数となるが、これを $W[j]$ と定めておく ($\mathcal{W}[y]$ との混同に注意)。するとこれを各補助変数 j で汎関数微分することで、

$$\begin{aligned} \langle R(t) \rangle &= \left. \frac{\delta W[j]}{\delta j_1(t)} \right|_{j=0}, \quad \langle \tilde{R}(t) \rangle = \left. \frac{\delta W[j]}{\delta j_2(t)} \right|_{j=0}, \quad \langle Q(t, s) \rangle = \left. \frac{\delta W[j]}{\delta j_3(t, s)} \right|_{j=0}, \quad \langle \tilde{Q}(t, s) \rangle = \left. \frac{\delta W[j]}{\delta j_4(t, s)} \right|_{j=0} \\ W[j] &:= \ln Z[j] \end{aligned}$$

が得られる。ここで $\langle \cdot \rangle$ は、クエンチ平均かつ試行平均を意味するが、自己平均性から試行平均のみとしても同じである。なお、このように $\ln Z[j]$ を j で汎関数微分することから、モーメント生成汎関数 $Z[j]$ を定義する際に Eq.3 の定数倍 $1/A$ は無視しても良く、以降ではそのようにする。つまり、

$$\begin{aligned} W[j] &:= \ln Z[j] \\ Z[j] &:= \int_y \exp \left(N \left(R^\top \tilde{R} + Q^\top \tilde{Q} + \mathcal{W}[y] \right) + j^\top y \right) \\ \mathcal{W}[y] &:= \ln \int_{\mathbf{x}, \tilde{\mathbf{x}}} \exp \left(S_0[x, \tilde{x}] - \tilde{x}^\top R + \frac{1}{2} \tilde{x}^\top Q \tilde{x} - \bar{g} \phi^\top \tilde{R} - g^2 \phi^\top \tilde{Q} \phi \right) \\ \implies \langle R(t) \rangle &= \left. \frac{\delta W[j]}{\delta j_1(t)} \right|_{j=0}, \quad \langle \tilde{R}(t) \rangle = \left. \frac{\delta W[j]}{\delta j_2(t)} \right|_{j=0}, \quad \langle Q(t, s) \rangle = \left. \frac{\delta W[j]}{\delta j_3(t, s)} \right|_{j=0}, \quad \langle \tilde{Q}(t, s) \rangle = \left. \frac{\delta W[j]}{\delta j_4(t, s)} \right|_{j=0} \end{aligned} \quad (4)$$

となる。ここまでは、任意のニューロン数 N に対して厳密に成り立つ結果である。

(2) 鞍点近似

$W[j]$ を解析的に厳密に計算することはできないが、 $Z[j]$ の \exp の肩にニューロン数 $N \gg 1$ に比例する項が存在することに着目すると、鞍点近似が使えらる。つまり、鞍点を $y_* = \{R_*, \tilde{R}_*, Q_*, \tilde{Q}_*\}$ とすると、

$$W[j] = \ln Z[j] \approx N \left(R_*^\top \tilde{R}_* + Q_*^\top \tilde{Q}_* + \mathcal{W}[y_*] \right) + j^\top y_* \quad (5)$$

となり、これを用いると Eq.4 より、

$$\langle R(t) \rangle \approx R_*(t), \quad \langle \tilde{R}(t) \rangle \approx \tilde{R}_*(t), \quad \langle Q(t, s) \rangle \approx Q_*(t, s), \quad \langle \tilde{Q}(t, s) \rangle \approx \tilde{Q}_*(t, s) \quad (6)$$

つまり、オーダーパラメータ y の平均は、鞍点に等しくなる。

鞍点は、鞍点条件 $\frac{\delta}{\delta y} \left(R^\top \tilde{R} + Q^\top \tilde{Q} + W[y] \right) = 0$ により得ることができ、これを計算すると、

$$R_*(t) = \bar{g} \langle \phi(t) \rangle_{\mathcal{L}}, \quad \tilde{R}_*(t) = \langle \tilde{x}(t) \rangle_{\mathcal{L}}, \quad Q_*(t, s) = g^2 \langle \phi(t) \phi(s) \rangle_{\mathcal{L}}, \quad \tilde{Q}_*(t, s) = g^2 \langle \tilde{x}(t) \tilde{x}(s) \rangle_{\mathcal{L}} \quad (7)$$

となる。なお、

$$\langle a \rangle_{\mathcal{L}} = \frac{\int_{x, \tilde{x}} a \exp \left(S_0[x, \tilde{x}] - \tilde{x}^\top R + \frac{1}{2} \tilde{x}^\top Q \tilde{x} - \bar{g} \phi^\top \tilde{R} - g^2 \phi^\top \tilde{Q} \phi \right)}{\int_{x, \tilde{x}} \exp \left(S_0[x, \tilde{x}] - \tilde{x}^\top R + \frac{1}{2} \tilde{x}^\top Q \tilde{x} - \bar{g} \phi^\top \tilde{R} - g^2 \phi^\top \tilde{Q} \phi \right)}$$

という表記を用いている。

ここで、 $\tilde{R}_* \equiv 0$ 、 $\tilde{Q}_* \equiv 0$ は self consistent な解になっている。実際、これらを仮定して Eq.5 に代入すると、

$$\begin{aligned} W[j] &\approx NW[R_*, 0, Q_*, 0] + j_1^\top R_* + j_3^\top Q_* \\ &= N \ln \int_{x, \tilde{x}} \exp \left(S_0[x, \tilde{x}] - \tilde{x}^\top R_* + \frac{1}{2} \tilde{x}^\top Q_* \tilde{x} \right) + j_1^\top R_* + j_3^\top Q_* \end{aligned}$$

となるが、これは、 N 個の各ニューロンが互いに独立に、 $(1 + \partial_t)x(t) = \eta(t)$ にしたがって時間発展していることを表している。ここで η は有色ガウスノイズで、 $\langle \eta(t) \rangle = R_*(t)$ 、 $\langle \eta(t)\eta(s) \rangle = Q_*(t, s) + D\delta(t-s)$ を満たしている ($\langle \cdot \rangle$ は η の実現値に関する平均だが、これは $\langle \cdot \rangle_{\mathcal{L}}$ に等しいことに注意)。外部入力 $\tilde{j}(t)$ を加えて、 $(1 + \partial_t)x(t) = \eta(t) + \tilde{j}(t)$ を考えると、そのモーメント生成汎関数は、

$$Z[j, \tilde{j}] = \int_{x, \tilde{x}} \exp \left(S_0[x, \tilde{x}] - \tilde{x}^\top (R_* + \tilde{j}) + \frac{1}{2} \tilde{x}^\top Q_* \tilde{x} + j^\top x \right)$$

となるから (前述の通り定数倍を無視していることに注意)、これを \tilde{j} で汎関数微分することにより、

$$\langle \tilde{x}(t) \rangle = \langle \tilde{x}(t) \rangle_{\mathcal{L}} = \left. \frac{\delta Z[j, \tilde{j}]}{\delta \tilde{j}(t)} \right|_{j=\tilde{j}=0}, \quad \langle \tilde{x}(t)\tilde{x}(s) \rangle = \langle \tilde{x}(t)\tilde{x}(s) \rangle_{\mathcal{L}} = \left. \frac{\delta^2 Z[j, \tilde{j}]}{\delta \tilde{j}(t)\delta \tilde{j}(s)} \right|_{j=\tilde{j}=0}$$

となる。生成汎関数の定義より $Z[j=0, \tilde{j}]$ は \tilde{j} によらず定数であるから直ちに、 $\langle \tilde{x}(t) \rangle_{\mathcal{L}} = \langle \tilde{x}(t)\tilde{x}(s) \rangle_{\mathcal{L}} = 0$ と分かる。よって、Eq.7 より、 $\tilde{R}_* = \tilde{Q}_* \equiv 0$ となり、当初の仮定と無矛盾である。

なお、 \tilde{R}_* 、 \tilde{Q}_* が非ゼロな解を持つ場合は、因果性の破綻、つまり、ある時刻に RNN に加えた摂動が、その時刻よりも前の活動に影響を与えるようなことが生じる (時間逆向きについて非ゼロな応答関数が導かれる) ため、非物理的な解になるらしい [7]。

(3) 動的平均場方程式の解析

前節で得られた結果を再度まとめる。大自由度極限 $N \rightarrow \infty$ では、RNN の各ニューロン x は、互いに独立に、以下の確率微分方程式にしたがって時間発展するとみなすことができる：

$$(1 + \partial_t)x(t) = \eta(t), \quad \eta \sim \mathcal{GP}(\bar{g} \langle \phi(t) \rangle, g^2 \langle \phi(t)\phi(s) \rangle + D\delta(t-s)) \quad (8)$$

ここで $\eta \sim \mathcal{GP}(R_*, Q_*)$ は、 η がガウス過程であり、 $\langle \eta(t) \rangle = \bar{g} \langle \phi(t) \rangle$ 、 $\langle \eta(t)\eta(s) \rangle = g^2 \langle \phi(t)\phi(s) \rangle + D\delta(t-s)$ であることを意味している。なお $\langle \cdot \rangle$ はガウス過程 η の実現値に関する平均である。 x がガウス過程 η で決まり、 η の従う確率分布は x に依存して決まるため、閉じた関係式となっており、Eq.8 を動的平均場方程式と呼ぶ。

定常状態では $\langle \eta(t) \rangle$ や $\langle \eta(t)\eta(t+\tau) \rangle$ は時刻 t によらず一定値を取るとみなせるので、 $\langle \eta(t) \rangle := R$ 、 $\langle \eta(t)\eta(t+\tau) \rangle := Q(\tau)$ とおける。 $\mu := \langle x(t) \rangle$ 、 $C(\tau) := \langle x(t)x(t+\tau) \rangle$ とし、 $x(t)$ も $\eta(t)$ と同様にガウス過程に従うことに着目すれば、式 (8) の 2 つ目の関係式より、

$$\begin{aligned} R &= \bar{g} \langle \phi(x) \rangle_{x \sim \mathcal{N}(x|\mu, C(0))}, \quad Q(\tau) = g^2 \langle \phi(x)\phi(y) \rangle_{(x,y) \sim \mathcal{N}(\mathbf{m}, \Sigma)} + D\delta(\tau) \\ \mathbf{m} &:= (\mu, \mu)^\top, \quad \Sigma := \begin{pmatrix} C(0) & C(\tau) \\ C(\tau) & C(0) \end{pmatrix} \end{aligned}$$

が得られる。つまり $(\mu, C(\tau))$ が分かれば $(R, Q(\tau))$ が分かる。逆に $(R, Q(\tau)) \rightarrow (\mu, C(\tau))$ の関係は、 $(1+\partial_t)x(t) = \eta(t)$ から得られる。まず両辺の統計平均を取ると、定常性から $(1+\partial_t)\mu = \mu = R$ 。同様に、 $(1+\partial_t)(1+\partial_s)x(t)x(s) = \eta(t)\eta(s)$ の統計平均を取ること、 $(1 - \partial_\tau^2)C(\tau) = Q(\tau)$ となる。

以上をまとめると、定常状態における動的な平均場方程式は、

$$(1 - \partial_\tau^2)C(\tau) = Q(\tau), \quad R = \bar{g} \langle \phi(x) \rangle_{x \sim \mathcal{N}(x|R, C(0))}, \quad Q(\tau) = g^2 \langle \phi(x)\phi(y) \rangle_{(x,y) \sim \mathcal{N}(m, \Sigma)} + D\delta(\tau) \quad (9)$$

というオーダーパラメータに関する微分方程式に帰着できる。

これを解けばオーダーパラメータの統計的性質が得られるが、一般には解析的に陽に解くことはできず、数値計算を要する。ただ定性的な議論によって相転移の大まかな様相は分かる。まず、 $R = 0$ は Eq.9 の第 2 式の固定点である。 $R \ll 1$ でこの方程式を線形近似すると、 $R = \bar{g}R \langle \phi'(y) \rangle_{y \sim \mathcal{N}(0, C(0))} + O(R^3)$ となり、このことから $R = 0$ は $\bar{g} \langle \phi'(y) \rangle_{y \sim \mathcal{N}(0, C(0))}$ が 1 を超えると不安定化することが分かる。 $R = 0$ のときには、Eq.9 の第 1, 3 式は $\bar{g} = 0$ の動的な平均場方程式になるので、先行研究 [1, 2] の結果をそのまま使える。この結果によれば、特に $D = 0$ のときには、 $g < 1$ で $C \equiv 0$ 、 $g > 1$ で C は非ゼロの関数になる。 $C \equiv 0$ なら、 $\bar{g} \langle \phi'(y) \rangle_{y \sim \mathcal{N}(0, C(0))} = \bar{g}$ ($\because \phi'(0) = 1$) であるから、ちょうど $\bar{g} = 1$ で $R = 0$ が不安定化する。一方、 $g > 1$ では $C(0) > 0$ となるから、 $\langle \phi'(y) \rangle_{y \sim \mathcal{N}(0, C(0))} < 1$ ($\because \phi''(x) < 0$) となり、固定点が不安定化する \bar{g} は 1 よりも大きくなる。この分岐点は g に依存するので、 $\bar{g}^*(g)$ と表しておく。

つまり、 $D = 0$ では、 $g < 1$ かつ $\bar{g} < 1$ ならば $R = 0$ かつ $Q = 0$ となるが、これは全てのニューロンが活動しない、つまり自明固定点 $x = 0$ を意味する。 $g < 1$ かつ $\bar{g} > 1$ では $R > 0$ となるが、この時、 x が新たな非自明固定点になるのが、カオスになるのかは $Q(\tau)$ が非ゼロ解になるかで決まる。 $g > 1$ かつ $\bar{g} < \bar{g}^*(g)$ のときは、 $R = 0$ かつ $Q > 0$ となるので、ダイナミクスは $x = 0$ 周りを揺らぐカオスとなる。一方、 $g > 1$ かつ $\bar{g} > \bar{g}^*(g)$ では、 $R > 0$ となるが、実は g が比較的小さければ $Q > 0$ 、大きいと $Q = 0$ となる。つまり、前者は非自明固定点 \bar{x} 周りを揺らぐカオス、後者は非自明固定点 x で静止した活動となる。実際の相図は以下になる (文献 [5] Fig.2 をそのまま掲載)。赤色は R の大きさ、グレーは $Q(0)$ の大きさを表している。

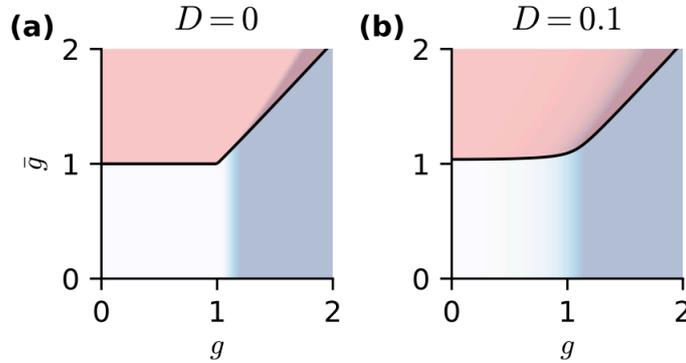


Figure 2. Mean-field phase diagram spanned by \bar{g} and g for (a) the noiseless case ($D = 0$) and (b) noise-driven dynamics ($D = 0.1$). The red shading quantifies the absolute population activity $|R|$, which is the order parameter for ferromagnetic activity and the gray shading quantifies the dynamic variability Q , which for $D = 0$ is the order parameter indicating the onset of chaotic activity. The black curves show where these values become nonzero. The dynamic variability Q does not vanish in the presence of noise.

図 2

なお、ここでは $Q > 0$ を以ってダイナミクスはカオスと言ったが、厳密には最大 Lyapunov 指数が正であることを確認する必要がある。しかし、同じ RNN を 2 つ用意して (レプリカ)、この 2 体系に対して動的な平均場理論を適用すれば、最大 Lyapunov 指数を解析的に求めることができる [2, 8]。

4 生成汎函数の経路積分表示を用いた手法の有用性

前章では、生成汎函数の経路積分表示を経由して動的平均場方程式 (Eq.8) を求めたが、本モデルに限れば、実はこれを素朴な議論で導出できる。モデル Eq.1 の右辺 第1項 $\sum_j^N J_{ij}\phi_j(t)$ は、 $N \gg 1$ 個の独立な確率変数 $J_{ij}\phi_j$ の和であるから、中心極限定理より漸近的にガウス分布に従い、平均は $\langle \sum_j^N J_{ij}\phi_j(t) \rangle_J = \sum_j^N \langle J_{ij} \rangle_J \langle \phi_j(t) \rangle_J = \bar{g} \langle \phi(t) \rangle$ 、自己相関関数は $\langle \sum_j^N J_{ij}\phi_j(t) \sum_k^N J_{ik}\phi_k(s) \rangle_J = \sum_{j,k} \langle J_{ij}J_{ik} \rangle_J \langle \phi(t)\phi(s) \rangle_J = g^2 \langle \phi(s)\phi(t) \rangle$ と分かり、Eq.8 が導かれる。ここで、 J_{ij} と ϕ_j の相関は $N \gg 1$ では無視できるとみなしていることに注意。

それなら、生成汎函数を経路積分表示してから鞍点近似をするという、一連の複雑な計算はそもそも必要ないように感じてしまう。本手法の提案者である Sompolinsky らは、経路積分表示のメリットとして、鞍点の安定性を見ることで、動的平均場近似で得られた解が安定であるかどうかを調べることができると述べており、文献 [8] では実際にこれを議論している。しかし、平均場近似解の安定性は（重要ではあるが）マニアックな話題であり、経路積分表示の恩恵を十分に感じる例ではない。

本章では、経路積分表示の有用性が分かる問題を 2 つ紹介し、その概要を述べる。

(1) シナプス結合に部分的な対称性がある RNN の動的平均場理論

Eq.1 のモデルでは、全てのシナプス重み J_{ij} は互いに独立としていたが、 J_{ij} と J_{ji} が互いに相関係数 $\tau \in [-1, 1]$ で相関しているようなモデル [9] を考える。 $\tau = 0$ では元のモデルに帰着される。このモデルでは、直感的に動的平均場方程式を導出することは困難となる。理由は次の通りである。 $\tau \neq 0$ のときには、 J_{ij} と J_{ji} が正に相関しているため、ニューロン $j \rightarrow$ ニューロン $i \rightarrow$ ニューロン j というルートを介して活動にフィードバックがかかる。 $|\tau|$ が大きいほどこの効果が強くなることから、 J_{ij} と ϕ_j の間には、たとえ $N \rightarrow \infty$ であっても無視できない相関が生じる。そのため上で計算したように素朴に $\sum_j^N J_{ij}\phi_j(t)$ の平均や自己相関関数を得ることができないのである。

したがって、このモデルでは前章のように経路積分表示を経由しないと動的平均場方程式が得られない。前章と同じ手順を踏んで計算を進めると、 $\bar{g} = 0$ の場合、動的平均場方程式は、

$$(1 + \partial_t)x(t) = \tau g^2 \int R(t, s)\phi(s)ds + \eta(t)$$

となる [9]。ここで $R(t, s)$ は response function、 $\eta(t)$ は従来モデルと同様のガウス過程を表す。右辺第1項は、前述した、結合相手のニューロンを介した自身の活動のフィードバックを反映している。この項を素朴な直感的議論で得ることは至難の業だろう。

なお、Cavity 法という統計物理で用いられる手法を適用すれば、経路積分表示を用いなくてもこの動的平均場方程式を導出することができる [10]。Cavity 法の方が物理的直感が効きやすく理解しやすい上に、経路積分表示に比べて少ない計算量で動的平均場方程式が得られるため、動的平均場方程式を求めるだけなら cavity 法で十分という説はある。

(2) 有限サイズ修正やニューロン間相関の計算

前章では生成汎函数 (Eq.3) を鞍点近似してオーダーパラメータ R, Q の統計平均を求めた。これは $N \rightarrow \infty$ では R, Q の揺らぎを無視することができるため厳密な結果になるが、有限サイズでは揺らぎが生じ、さらにその揺らぎの影響を受けて R, Q の統計平均も、鞍点近似解から $O(1/N)$ の修正を受ける。このような有限サイズ効果を $O(1/N)$ の項までちゃんと計算するために、Eq.3 のモーメント生成汎函数 Z に、オーダーパラメータの 2 次の項に関する source term k を追加で導入し、 $Z[j, k] = \int_y \exp(NW[y] + j^\top y + \frac{1}{2}y^\top k^\top y)$ とする。なお

$$y = (R, \tilde{R}, Q, \tilde{Q})^\top, \quad j = (j_1, j_2, j_3, j_4)^\top, \quad k \text{ は関数を成分にもつ } 4 \times 4 \text{ 行列であり, } j = 0 \text{ かつ } k = K := \begin{pmatrix} 0100 \\ 1000 \\ 0001 \\ 0010 \end{pmatrix}$$

で $\exp(NW[y] + j^\top y + \frac{1}{2}y^\top k^\top y)$ は元の確率測度に帰着される。 $Z[j, k]$ の 2 変数ルジャンドル変換を

$$\Gamma[\alpha_1, \alpha_2] := \sup_{\alpha_1, \alpha_2} j^\top \alpha_1 + k^\top \alpha_2 - \ln Z[j, k]$$

として、いわゆる”状態方程式”[6] $\frac{\delta \Gamma}{\delta \alpha_1} = 0, \frac{\delta \Gamma}{\delta \alpha_2} = \frac{1}{2}NK$ を解けば、有限サイズ効果を反映した R, Q の統計平均および自己相関が得られる。もちろん Γ を厳密に求めることはできないため、この状態方程式を厳密に解くことはできないが、 Γ が場の量子論における有効作用と形式的に対応することに着目すると、 $\Gamma[j, k]$ を loop 展開により近似計算したものをいれれば解析的に解ける。leading order で展開を切った Γ (tree-level 近似) を用いると、実は、状態方程式は鞍点方程式と同じものになる。つまり、平均場近似解に帰着される。一方、 Γ を subleading order まで展開したもの (1-loop 近似) を用いて状態方程式を解くと、subleading order までの解、つまり有限サイズ効果を組み込んだ解が得られる。また、ニューロン間相関 $\frac{1}{N^2} \sum_{i,j(i \neq j)} \phi_i(t)\phi_j(s)$ は、簡単な計算により $\langle R(t)R(s) \rangle - \frac{1}{N} \langle Q(t, s) \rangle$ と等しいと分かり、これは状態方程式の解 $\langle Q(t, s) \rangle, \langle R(t)R(s) \rangle$ を用いれば求められる。詳細な計算は非常に複雑なので、元論文 [5] を見てほしい。

この手法を用いて計算すると、自明固定点から非自明固定点への分岐点 $\bar{g} = 1$ ($g < 1$) における $\langle Q(\tau) \rangle$ は、動的平均場近似解から $O(1/N)$ の修正を受けるが、この修正項には減衰が極めて遅い自己相関関数 $\langle R(t)R(t+\tau) \rangle$ ($\bar{g} \nearrow 1$ で decay time は発散) が含まれるため、 $\langle Q(\tau) \rangle$ もその影響を受けて減衰が非常に遅い関数になる (下図 (c) 参照; 文献 [5]Fig.1 をそのまま掲載)。

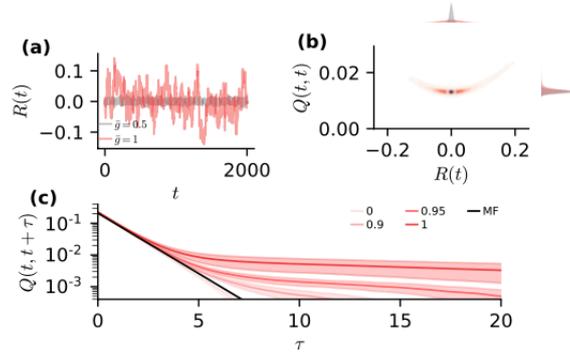


Figure 1. (a) Population-averaged activity $R(t)$ for $\bar{g} = 0.5$ (gray) and $\bar{g} = 1$ (red). (b) Auxiliary fields Q and R , proportional to population-averaged output autocorrelation and activity, respectively, binned for each point in time. (c) Time-lagged, population-averaged, stationary autocorrelation $Q(t, t + \tau)$ simulated for different values of \bar{g} (shades of red) and mean field prediction (black) plotted logarithmically. Remaining network parameters: $\phi(x) = \text{erf}(\frac{\sqrt{x}}{2}x)$, $N = 1000$, $g = 0.5$, and $D = 0.1$.

図 3

なお、ここでも実は、経路積分表示の代わりに、Cavity 法 (two-site cavity 法 [11]) によってニューロン間相関を求めることができる。この場合は前節と異なり、Cavity 法でも経路積分表示の手法でも、どちらも計算は大変になるのだが、Cavity 法の方がやはり物理的直感は効きやすい。

ということで、Cavity 法を使えば、必ずしも経路積分表示を用いる必要はない。ではどちらを使えば良いのだろうか。これはケースバイケースとしか言えないだろうが、両者のメリットとデメリットについて最後に言及しておく。経路積分表示は source term j, k や補助場 y など様々な補助変数が出てくるため、物理的なイメージが掴みづらく、途中で何をしているのか迷子に陥りやすい。その一方、Cavity 法は物理的な直感に根ざした手法であるため、理解しやすく、また場合によっては経路積分表示の手法よりも計算が簡単になることがある。しかし、各項のオーダーをその都度気にしながら計算を進めなくてはならず、ad hoc な対応が求められる。その反面、経路積分表示を経由する手法は、システマティックに計算を進められるため、経路積分表示→鞍点近似や loop 展開という一連の処方箋に一度慣れてしまえば、こちらの方が使いやすいかもれない。

Cavity 法の詳細に興味ある人は文献 [10, 11] を読んでください。

参考文献

- [1] H. Sompolinsky, A. Crisanti, and H. J. Sommers. Chaos in random neural networks. *Phys. Rev. Lett.*, 61:259–262, Jul 1988.
- [2] Jannis Schuecker, Sven Goedeke, and Moritz Helias. Optimal sequence memory in driven random networks. *Phys. Rev. X*, 8:041029, Nov 2018.
- [3] Kanaka Rajan and L. F. Abbott. Eigenvalue spectra of random matrices for neural networks. *Phys. Rev. Lett.*, 97:188104, Nov 2006.
- [4] Terence Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Relat. Fields*, 155:231–263, 2013.
- [5] Michael Dick, Alexander van Meegen, and Moritz Helias. Linking network and neuron-level correlations by renormalized field theory. *arXiv*, 2024.
- [6] Moritz Helias and Dahmen David. *Statistical Field Theory for Neural Networks*. Springer, 2020.
- [7] H. Sompolinsky and Annette Zippelius. Relaxational dynamics of the edwards-anderson model and the mean-field theory of spin-glasses. *Phys. Rev. B*, 25:6860–6875, Jun 1982.
- [8] A. Crisanti and H. Sompolinsky. Path integral approach to random neural networks. *Phys. Rev. E*, 98:062120, Dec 2018.
- [9] Daniel Martí, Nicolas Brunel, and Srdjan Ostojic. Correlations between synapses in pairs of neurons slow down dynamics in randomly connected neural networks. *Phys. Rev. E*, 97:062314, Jun 2018.
- [10] Wenxuan Zou and Haiping Huang. Introduction to dynamical mean-field theory of randomly connected neural networks with bidirectionally correlated couplings. *arXiv*, 2023.
- [11] David G. Clark, L. F. Abbott, and Ashok Litwin-Kumar. Dimension of activity in random neural networks. *Phys. Rev. Lett.*, 131:118401, Sep 2023.